

ISSN 0027-8084

УДК 808.61./62

**ИНСТИТУТ ЗА СРПСКИ ЈЕЗИК САНУ**

---

# **НАШ ЈЕЗИК**

**LI/2**

**БЕОГРАД**

**2020**

# НАШ ЈЕЗИК

ЛІ СВ. 2

## УРЕЂИВАЧКИ ОДБОР

др Рајна ДРАГИЋЕВИЋ, др Јелена ЈОВАНОВИЋ СИМИЋ,  
др Весна ЛОМПАР, др Драгана МРШЕВИЋ РАДОВИЋ,  
др Марина НИКОЛИЋ, др Мирослав НИКОЛИЋ,  
др Ђорђе ОТАШЕВИЋ, др Милица РАДОВИЋ ТЕШИЋ,  
др Стана РИСТИЋ, др Марина СПАСОЈЕВИЋ,  
др Срето ТАНАСИЋ

Уредник  
др СРЕТО ТАНАСИЋ

Секретар  
др Марина Спасојевић

Рецензенти  
др Ђорђе Оташевић, др Милица Радовић Тешић,  
др Марина Спасојевић, др Срето Танасић

Према мишљењу Министарства за науку, технологију и развој Републике Србије  
бр. 413-00-1715/2001-01 ова публикација је од посебног интереса за науку.

НАШ ЈЕЗИК, часопис за српски књижевни језик, покренуло је 1932. г. Лингвистичко друштво у Београду, на челу са Александром Белићем, који је уредио седам књига предратне серије, као и девет књига и први двоброј десете књиге у новој, послератној серији часописа. Од 1960. до 1975. г. (књ. X/3–XXI/1–2) уредник часописа био је Михаило Стевановић, а у периоду од 1975. до 1996. г. (књ. XXI/3–XXX) Митар Пешикан. Часопис је од 1996. до 2005. г. (књ. XXXI–XXXVI) уређивао Мирослав Николић.

Наш језик излази у две годишње свеске, које чине једну књигу. Издаје га и прима рукописе Институт за српски језик САНУ, Кнез Михаилова 36, ел. пошта: [isj@isj.sanu.ac.rs](mailto:isj@isj.sanu.ac.rs), [nasjekisjsanu@gmail.com](mailto:nasjekisjsanu@gmail.com).

Штампа: Службени гласник, Београд

CIP – Каталогизација у публикацији  
Народна библиотека Србије, Београд

81

**НАШ језик** / уредник Срето Танасић. – Год. 1, бр. 1 (1932/1933) – год. 8, бр. 1 (1941); н. с. књ. 1, св. 1/2 (1949/1950)–. – Београд: Институт за српски језик САНУ, 1933–1941; 1949– (Београд : Службени гласник). – 23 cm

Полугодишње

ISSN 0027-8084 = Наш језик

COBISS.SR-ID 615951

# НАШ ЈЕЗИК

Књига LI

Св. 2 (2020)

## САДРЖАЈ

### Ч л а н ц и

Реконструкција сугласника у критичком издању Његошевог спјева <i>Шћепан Мали</i> (РАДМИЛО Н. МАРОЈЕВИЋ) .....	1–37
Откривање неконсистентности у обради вишеструке еквиваленције: уводна размишљања (ДАНКО М. ШИПКА) .....	39–49
Идеофонски предикати у српском језику: основне карактеристике (ТАЊА Ч. МИЛОСАВЉЕВИЋ) .....	51–60

### П р и к а з и

Вуков <i>Српски рјечник и 200 година савременог српског језика</i> (ДРАГАНА Ј. ЦВИЈОВИЋ) .....	61–66
Рајна Драгићевић, <i>Грамањика у огледалу семантике</i> (БОЈАНА Д. ТОДИЋ САНКОВИЋ) .....	67–70
Веселина Ђуркин, <i>Сложени зависни везници у српском језику</i> (и њихова функционално-стилска дистрибуција) (МАРИНА М. НИКОЛИЋ) .....	71–74
Милка Николић, <i>Поредбено-начинске конструкције у српском језику</i> (АНЂЕЛА Д. ВАСИЉЕВИЋ) .....	75–82

### Х р о н и к а

Хроника Међународног научног скупа „Лексикографија и лексикологија у светлу актуелних проблема“ (28–30. октобар 2020) (СЛОБОДАН Б. НОВОКМЕТ) .....	83–95
----------------------------------------------------------------------------------------------------------------------------------------------------------	-------

## **Одлуке Одбора за стандардизацију српског језика**

Терминолошке синтагме које конституише именица <i>менаџмент</i> . . .	97–99
Писање назива једног вируса: <i>вирус корона</i> . . . . .	101–104

## **In memoriam**

Радоје Симић (ВИКТОР Д. САВИЋ) . . . . .	105–112
Владимир П. Гутков (МАРИНА Љ. СПАСОЈЕВИЋ) . . . . .	113–116

УДК 811.163.41'374  
811.111'374

Оригинални научни рад  
Примљено 20. септембра. г.  
Прихваћено 31. децембра 2020. г.

ДАНКО М. ШИПКА\*

(Државни универзитет Аризоне,  
Одсјек за њемачки, румунски и словенске језике;  
Универзитет Колумбија,  
Одсјек за словенске језике)\*\*

## ОТКРИВАЊЕ НЕКОНСИСТЕНТНОСТИ У ОБРАДИ ВИШЕСТРУКЕ ЕКВИВАЛЕНЦИЈЕ: УВОДНА РАЗМИШЉАЊА

У раду се износе уводна размишљања о састојцима система за откривање неконсистентности у обради вишеструке еквиваленције у двојезичним рјечницима. Након што су релевантни појмови објашњени, у раду се идентификују и опримјерују типови неконсистентности у обради вишеструке еквиваленције. У слједећем дијелу рада дају се начела на основу којих би такав систем за откривање неконсистентности био сачињен, а затим се описују обрасци претраживања рјечничког текста у циљу идентификације неконсистентности, што је заправо срж таквог једног замишљеног система. У завршном дијелу рада изводе се општији закључци на основу свега представљеног и дају назнаке могућег рада на овом замишљеном систему. Овај рад садржи уводна размишљања и замишљен је као први корак у развоју оваквог система подршке лексикографима који раде на двојезичним рјечницима. Нека од рјешења могу се у модификованом облику примјенити и на једнојезичне рјечнике.

*Кључне ријечи:* вишеструка еквиваленција, лексички анисоморфизам, лексикографска обрада, провјера конистенције, двојезична лексикографија.

### Појмовна мапа

У овом раду разматрају се ситуације неконсистентности лексикографске обраде који произлазе из вишеструке еквиваленције, ситуације

---

\* danko.sipka@asu.edu

\*\* Дио грађе за овај рад прикупљен је током истраживачког боравка у Минхену 2011. године, који је финансирала фондација „Александар фон Хумболт“.

гдје једној ријечи у изворишном језику у рјечничком чланку одговарају двије или више ријечи одредишног језика. Вишеструка еквиваленција врста је лексичког анисоморфизма, појаве да између два језика која се на било који начин упарују не постоји лексичка еквиваленција један на један а без остатка (а такве пуне еквиваленције у већини случајева заправо и нема). Лексички анисоморфизам релевантан је у различитим облицима дјеловања. Унутар ове појаве, на значај вишеструке еквиваленције вишеструко се упућивали у металексикографији (нпр. Згуста 1971; Ал Касими 1977; Јонг–Пенг 2007, да поменемо само најзначајније приручнике). Слично је и у филозофији природног језика (нпр. Фреге 1982; Патнам 1975), лингвистичкој антропологији (нпр. Годард 2012), транслатологији (нпр. Пим 2009), те настави страног језика (нпр. Крамш 1993).

Најједноставнији начин да се класификују случајеви лексичког анисоморфизма, како смо показали у Шипка 2015, јесте да се преброји број еквивалената, и ту су могуће три ситуације. Прва је нулта еквиваленција, гдје ријеч у изворишном језику нема еквивалента у одредишном. На примјер наше ријечи *слава* или *кољиво* неће имати еквиваленте у многим језицима. Друга је ситуација дјеломичне еквиваленције, гдје постоји еквивалент у одредишном језику, али се он у неком битном аспект у функционисања разликује од оног у изворишном. На примјер, наше *значка* има еквивалент у енглеском *lapel pin*, али је фреквенција и препознатљивост овог еквивалента знатно мања него код нас, па нема метафорских проширења као у оном Радовићевом: „Све је пошло наопачке за врапце и мачке кад је јесен окачила своје жуте значке.“ Коначно, ситуација којом се овдје бавимо је вишеструка еквиваленција, гдје једној ријечи изворишног језика одговарају двије или више ријечи одредишног. Рецимо, наша ријеч *рачун* има веома много енглеских еквивалената: *calculus* (то је интегрални и диференцијални рачун), *account* (банковни рачун), *bill* (рачун за комуналије), *check* (за цех у ресторану), *receipt* (у продавници) итд.

Лексикографска обрада поменутих типова анисоморфизма знатно се разликује. Код нулте еквиваленције мора се објаснити о чему се ради. Ту, умјесто у двојезичним рјечницима уобичајених еквивалената, мора да се пружи дефиниција типа оних у једнојезичним рјечницима, само, наравно, на одредишном језику. Код дјеломичне еквиваленције треба упутити на оне особине по којима се еквиваленти у неком кључном погледу разликују. Коначно, обрада вишеструке еквиваленције треба да се заснива на раздвајању еквивалената.

Наравно, то како ће се и да ли ће се у рјечнику обрађивати лексички анисоморфизам зависи од потреба корисника и типа рјечника. Тако, ако у поменутом примјеру *рачун* имамо српско-енглески рјечник чији је једини задатак да помаже говорницима енглеског језика у интерпретативним

вјештинама (рецимо, разумијевању писаног текста на српском), онда је довољно представити еквиваленте без додатних напомена. Ако пак речени српско-енглески рјечник помаже говорницима српског у комуникативним вјештинама (рецимо, говорењу на енглеском), онда се том говорнику мора рећи кад да користи који еквивалент.

Под неконсистентношћу у обради вишеструке еквиваленције (а то је тема овог рада) подразумевамо неадекватно раздвајање еквивалената у одредишном језику. Разлози због којих је обрада неконсистентна различити су, а претраживање неконсистенције различито је код сваког појединог типа неконсистентности. Тако ћемо у другом дијелу рада представити типове неконсистентности, у трећем критеријуме које систем за претраживање неконсистентности треба да задовољи, а у четвртном предложити обрасце претраживања, односно показати кад су уопште могући и са којим степеном ефикасности. Овај рад садржи уводна размишљања и замишљен је као први корак у могућем развоју система за откривање недоследности у обради вишеструке еквиваленције у двојезичним рјечницима. Нека од рјешења дају се у модификованом облику примијенити на неконсистентности у разграничавању значења (дакле, код вишеструких дефиниција) у једнојезичним рјечницима.

### Типови неконсистентности

У овом дијелу представићемо и опримјерити типове неконсистентности у обради вишеструке еквиваленције. Примјери које користимо су хипотетички. Свјесно је одабрана стратегија да се не користе примјери из конкретних рјечника. Ради се о томе да се у сваком рјечнику може наћи одређена доза неконсистентности, па би било некоректно давати примјере из постојећих рјечника (штетећи тако њиховом угледу) за оно што се фактички може наћи у било којем од њих.

Сама грађа прикупљена је у оквиру пројекта писања монографије о лексичком анисоморфизму (Шипка 2015), а речена грађа потиче из анализе различитих двојезичних рјечника језика из различитих географских области, а који представљају веома различите језичке породице.

Издвојени су сљедећи типови могуће неконсистентности у обради вишеструке еквиваленције:

1. Неодговарајућа стратегија
2. Нејасност
3. Недостатак објашњења или спецификације
4. Редунданција
5. Подспецификацијско преклапање



6. Надспецификацијске празнине
7. Погрешна квалификација
8. Недостатно опримјерење
9. Употреба неодговарајућег језика
10. Недосљедно раздвајање
11. Ненавођење еквивалената
12. Неодговарајућа интерпретација.

Размотримо сада сваки од наведених дванаест типова неконсистентности понаособ. **Неодговарајућа сјрапйегуја** јавља се у случајевима кад лексикограф примјењује стратегију која је примјеренија другом типу рјечника а не ономе на коме ради. На примјер, у српско-њемачком рјечнику дефиниција типа *зид*, а м *Wand, Mauer* сасвим је прихватљива у случају рецепцијског рјечника за њемачке кориснике, али не и у продукцијском рјечнику за српске кориснике, јер ови други овако неће знати кад да користе један, а кад други еквивалент. **Нејасносћ** видимо у хипотетичком примјеру из српско-енглеског рјечника *косац*, *сца* м 1. *agricultural worker*, 2. *a kind of bird*, 3. *a sort of insect*. Сва три значења су дефинисана прешироко, еквивалент јесте унутар дефиниције, али језички корисник са енглеског подручја неће знати о којој врсти пољодјелца, птице и инсекта се ради. **Недостатак објашњења или сјецификације** јавља се у случајевима кад се у одредишном језику користи вишезначна ријеч, нпр. у српско-енглеском рјечнику, то би било у примјеру: *извор*, а м *spring, source*. Први енглески еквивалент уз значење извора има и значења 'опруга', 'прољеће' и друга, па је неопходно да назначи о којем се значењу еквивалента ради. **Редунданција** се јавља када се умјесто кратких напомена чисто да би се еквиваленти разграничили користе енциклопедијске дефиниције. На примјер, уместо да се у српско-енглеском рјечнику каже *рука*, е ж *hand (шака), arm (цпо уд)* редунданција би била у случају да се напише *рука*, е ж *hand (горњи екстремитет који служи за дохватање предмета и управљање њима), arm (горњи уд тела, који обухвата делове између раменских и лакатних зглобова и изван до прстију)*, што беспотребно оптерећује корисника. **Подсјецификацијско йреклајање** јавља се тамо гдје је један еквивалент постављен прешироко тако да покрива други. На примјер, у јужним аустралијским водама живи биће унеколико слично морском коњицу које се на енглеском зове *sea dragon* (дословно: *морски змај*). У њемачком се само за мању врсту тог бића користи назив *Seedrache* (дословно: морски змај, алтернативно, та подврста се може звати и *kleine Fetzenfish*, дословно: мала риба-изрезак), док се већа подврста зове *große Fetzenfish* (дословно: велика риба-

-изрезак). Уколико би за ово друго у енглеско-њемачком рјечнику било наведено само *Fetzenfish* (дословно: риба-изрезак), онда би овај еквивалент покрио и онај други јер би био подспецифициран, па би тако дошло до преклапања. **Надспецификацијске њразнине** супротне су од подспецификацијских преклапања у смислу да је један од еквивалената преузак, па цијело поље ријечи изворишног језика није покривено еквивалентима из одредишног. Рецимо, цвијет *звончић* на енглеском је *bluebell*. Поље те ријечи покривају двије горњолужичке ријечи: *česnička* и *kitelnička*. Уколико би умјесто овог првог ваљаног еквивалента у енглеско-горњолужичком стајала подврста, рецимо *zwónčkata česnička*, онда би тај еквивалент био надспецифициран, па би настала празнина. **Погрешна квалификација** може се видјети у овом хипотетичком чланку српско-њемачког рјечника: *маџура, е ж Abitur, Abi (у Њемачкој), Matura (у Швајцарској), Maturität (у Аустрији)*. Прво, замијењене су Швајцарска и Аустрија, а од двије ријечи у употреби у Њемачкој прва је неутрална, а друга неформална, па се погрешна квалификација састоји у ненавођењу квалификатора неформалности. **Недостатак оџримјерење** јавља се код функционалних ријечи. Уколико рецимо у рјечнику гдје је српски циљни језик за еквивалент приједлога *до* дамо само примјер просторне употребе, нпр. *сједи до колеџинице*, онда корисник може помислити да временска употреба (нпр. *ради до њеџка*) не постоји. **Уџотреба неодговарајућег језика** види се у случајевима гдје се напомене за разграничење еквивалената одредишног језика дају у том језику а не у изворишном. На примјер, ако бисмо у српско-енглеском рјечнику ставили *меџар, џира м meter (1.09 yard), tape measure (flexible ruler)*, умјесто одговарајућег: *meter (мјерна јединица), tape measure (алаџ)* да би се српски корисник информисао кад користити који еквивалент. **Недостатак раздвајање** јавља се кад у неким случајевима вршимо разграничавање еквивалената, а у неким не. На примјер, у одредници *рачун, а м bill (за комуналије), receipt (за неџио шџо се куџи), check (у ресџорану), account, calculus, interest* – прва три су разграничена, а остали нису. **Ненавођење еквивалената** имали бисмо у случају да у примјеру *рачун* испустимо неки од еквивалената. Коначно, **неодговарајућа инџерџретација** била би у случају да успостављамо помоћу еквивалената семантичко проширење које у изворишном језику не постоји. На примјер, у курманџијској курдској ријечи *dirûnker* изгледа као да постоји веза између значења жетелац и кројач, али те везе заправо нема, па ако бисмо у курдско-српском рјечнику ставили: *жеџелац, (меџафорски) кројач*, то би био случај неодговарајуће интерпретације.

### Критеријуми претраживања

Сад кад су одређени типови неконсистентности, прећи ћемо на критеријуме које систем за њено претраживање у тексту рјечника треба да задовољи. Претраживање треба да се руководи сљедећим општим начелима. Прво, било каква рјешења морају бити изводива. Може се замислити некакав идеалан систем претраживања, али ако се он не може провести у конкретном контексту лексикографског рада, онда је такво рјешење бескорисно. Друго, ресурси потребни за претраживање нужно су комплексни. Ово произлази из чињенице да су и типови неконсистентности представљени у претходном дијелу изразито разнородни, па онда и претраживање полази од различитих ресурса потребних за њихову идентификацију. Треће, битна је обухватност. Циљ система претраживања је да се нађу или сви случајеви неконсистентности или, кад то није могуће, што већи њихов број. Коначно, битна је искључивост, да систем претраживања налази само случајеве неконсистентности, а не и оно што ту не спада (дакле, ваљану обраду). У идеалном случају, а видјећемо да су реалне могућности далеко од тога, систем треба да пронађе све случајеве неконсистентности у раздвајању вишеструких еквивалената и само њих.

Кад је у питању сам резултат претраживања, пожељна су три могућа исхода. Први је идентификовање нечега као недоследности, други етикетирање као могућег случаја недоследности за подробнију провјеру, а трећи пропуштање, дакле, неуврштавање одређене одреднице или њеног дијела у категорију недоследност.

У идеалном случају систем претраживања био би у јавном домену да би омогућио кориштење лексикографима са ограниченим средствима за извођење пројеката. Добро би било омогућити локалну употребу да би систем могли користити лексикографи са нестабилним доступом интернету. На крају, у идеалном случају систем би био прилагодив конкретном пројекту. Различити пројекти укључују различите језике, чији ресурси варирају од веома богатих до крајње сиромашних. Такође, платформе на којима лексикографи раде, па онда и степен рашчлањености и форматираности рјечничке структуре изразито варира од пројекта до пројекта.

### Обрасци претраживања

Прећи ћемо сада на конкретне обрасце претраживања које би систем заснован на принципима назначеним у наредном дијелу могао укључити. Обрасци ће бити представљени за сваки од типова неконсистентности.

Неодговарајућа стратегија податна је на једноставан образац претраживања, гдје се, рецимо, уколико се ради о продукцијском рјечнику, може да се тражи ниска знакова без заграда. На примјер уколико су дефиниције обројчене а сваки еквивалент одвојен тачком зарезом, образац претраживања тражио би број иза којег слиједи ниска знакова гдје нема заграда, што би исказано регуларним изразима (о којима више в. у Ват 2005) изгледало овако:  $\backslash d[^(<)+?[\backslash ;\backslash r\backslash n]$ . Успјешност оваквог обрасца претраживања зависи од тога колико је рјечнички текст добро структуриран и колико је то дослиједно рађено. Уколико се, рецимо, шиљате заграде користе за напомене којима се разграничавају еквиваленти, онда ће образац претраживања бити веома ефикасан, уколико се користе обле заграде, ниво прецизности ће опати јер се оне могу користити и за напомене друге врсте.

Нејасност би се откривала тражењем дословних низака, типа „врста“, „тип“, итд. (рецимо, гдје се каже врста птице, тип плеса). Резултат овог механизма претраживања свакако би требало да буде етикетирање случајева за даљу провјеру. Ствар је у томе да се ове ниске могу легитимно појавити као додатна објашњења и слично. Тако онда образац има висок ниво надоткривања (дакле, етикетирања и онога што није случај нејасности), али и подоткривања (јер нејасност може да буде присутна и без присуства ових дословних низака).

Недостатак објашњења или спецификације може се открити обрасцем који се заснива на подацима из једнојезичне лексикографије. О овоме смо већ писали у Шипка 2007. Рецимо, ако се ради о српско-енглеском рјечнику, случајеви гдје недостаје објашњење или спецификација могу се наћи помоћу општедоступних података из Ворднета (2010). Тако, ако имамо одредницу као *кран*, а *m crane*, њој недостаје спецификација да се у енглеском ради о значењу грађевинске машине, а не птице ждрала. Образац једноставно налази овакве ситуације гдје је ријеч која је у Ворднету дата као вишезначна и лексикографу представља обраду у Ворднету са знаком да изабере о којем се значењу ради:  $|crane|large\ long-necked\ wading\ bird\ of\ marshes\ and\ plains\ in\ many\ parts\ of\ the\ world|lifts\ and\ moves\ heavy\ objects;\ lifting\ tackle\ is\ suspended\ from\ a\ pivoted\ boom\ that\ rotates\ around\ a\ vertical\ axis.$

Редунданција се може открити помоћу чисто формалних критеријума, рецимо, тражења низака од пет и више ријечи унутар обројчене дефиниције:  $\backslash d(\backslash b.+?\backslash b)\{5,\}$  те низака одвојених са три и више зареза или тачака зареза  $\backslash d(\backslash b.+?\backslash b[,:])\{3,\}$ .

Подспецификацијско преклапање и надспецификацијске празнине не дају се открити једноставним механизмима. За њихово откривање

биле би потребне изузетно прецизне семантичке мапе, па овај сегмент механизма откривања недоследности заправо није изводив.

Погрешна квалификација открива се обрасцима који се ослањају на податке из једнојезичних рјечника. Поређењем ауторитативног једнојезичног и двојезичног рјечника у питању биле би издвојене за провјеру ситуације гдје је: квалификатор присутан у једнојезичном рјечнику а није у двојезичном, квалификатор није присутан у једнојезичном рјечнику а јесте у двојезичном, те они гдје је етикета у два рјечника која се пореде различита.

Недостатно опримјерење слабо је податно на претраживање. Могу се претраживати примјери за провјеру. Рецимо, ако се примјери дају у облим или шиљатим заградама, могу се тражити кратке ниске иза којих су те заграде (у основи проблеми са погрешним опримјерењем десиће се кад нема додатних објашњења након еквивалента већ само примјер). Такав регуларни израз може бити  $d(b.+?b)\{1,2\} [(<].+$ . Наравно, ово претраживање одликује се високим степеном надоткривања јер ће само неки од издвојених случајева садржати недостатно опримјерење.

Употреба неодговарајућег језика открива се на основу листа за провјеру правописа даног језика. У случајевима језика гдје такве листе нису доступне, могу се израђивати ад хок стемери (о чему више в. у Кешел–Шипка 2008). Образац претраживања проналазио би све случајеве гдје на страни одредишног језика у заградама (или оном графичком облику који се користи за напомене разграничавања еквивалената) имамо облике који нису на листи облика ријечи изворишног језика.

Недоследно раздвајање заснива се на претраживању кратких ниски ријечи одвојених зарезима или тачкама зарезима, рецимо помоћу оваквог регуларног изрази  $d(b.+?b)\{1,2\} [;.]$ . Циљ је екстракција за провјеру, узевши у обзир да ће у многим издвојеним случајевима неће бити недоследног раздвајања.

Ненавођење еквивалената је слабо податно на претраживање, али ипак се може замислити процедура гдје се поређење број значења у једнојезичном рјечнику истог обима не поклапа са бројем еквивалената. Наравно, у великом броју случајева тај број и не треба да се поклапа, еквивалент није исто што и дефиниција значења у једнојезичном рјечнику, па ова процедура садржи висок ниво натпроналажења и потпроналажења и питање је да ли је уопште продуктивна.

Идентификација неодговарајуће интерпретације није изводива јер би морала да се ослања не само на прецизном семантичком мапирању него и на мапирању развоја значења.

## Закључак

На основу свега представљеног може се закључити слједеће. Прво, тип неконсистентности у великој мјери одређује изводивост, а онда и ефикасност образаца претраживања. Друго, изводивост и ефикасност образаца претраживања варира од оних који су неизводиви до оних који имају релативно висок ниво ефикасности. Треће, ниво формализованости рјечничког текста у великој мјери одређује ефикасност образаца претраживања. Рјечник гдје су сви елементи микроструктуре јасно разграничени, рецимо, у неком систем тагирања типа LMF (више у Франкопуло 2013) или TEI (више у Будин и др. 2012), и кодирани у XML-у (<http://www.w3.org/XML/>) свакако омогућује неупоредиво веће могућности претраживања од чистог текста, поготово ако у њему форматирање није досљедно. Међутим, како год да рјечнички текст изгледа са формалне стране, представљеним механизмима претраживања може се открити дио недосљедности, па тиме и побољшати квалитет рјечника.

Даљи рад на овој проблематици у идеалном случају водио би изградњи система за откривање недосљедности у обради вишеструке еквиваленције који би укључио све побројане образце претраживања и био флексибилан према степену рашчлањености и форматираниости рјечничког текста, као и према ресурсима који су доступни за дана два језика у двојезичном рјечнику. Наравно, у идеалном случају такав систем био би општедоступан. Дио рјешења био би релевантан и за једнојезичне рјечнике јер се неке неконсистентности, у нешто измијењеној форми, појављују и као проблем недосљедног разграничавања дефиниција значења ријечи у једнојезичном рјечнику.

## ЛИТЕРАТУРА

- Ал Касими 1977: Ali M. Al-Kasimi, *Linguistics and Bilingual Dictionaries*, Leiden: E. J. Brill.
- Будин и др. 2012: Gerhard Budin et al., Creating Lexical Resources in TEI P5 A Schema for Multi-purpose Digital Dictionaries, *Journal of the Text Encoding Initiative* 3, <http://jtei.revues.org/522>.
- Ват 2005: Andrew Watt, *Beginning Regular Expressions*, Birmingham: Wrox.
- Ворднет 2010: *WordNet. A Lexical Database for English*, Princeton: Princeton University, <http://wordnet.princeton.edu>.
- Годард 2012: Cliff Goddard, Semantic primes, semantic molecules, semantic templates: Key concepts in the NSM approach to lexical typology, *Linguistics* 50/3, 711–743.
- Зругра 1971: Ladislav Zgusta, *Manual of Lexicography*, The Hague: Mouton.

- Јонг–Пенг 2007: Heming Yong, Jing Peng, *Bilingual Lexicography from a Communicative Perspective*, Amsterdam: John Benjamins.
- Кешель–Шипка 2008: Владо Кешель, Данко Шипка, Приступ изградњи степера и лематизатора за језике с богатом флексијом и оскудним ресурсима заснован на обухватању суфикса, *ИНФОшкола* 9, 21–31.
- Крамш 1993: Kramsch, Claire *Context and culture in language teaching*, Oxford: Oxford University Press.
- Патнам 1975: Hilary Putnam, *Mind, Language and Reality. Philosophical Papers*, vol. 2. Cambridge: Cambridge University Press.
- Пим 2009: Anthony Pym, Natural and Directional Equivalence in Theories of Translation, in: *Metalanguage of Translation*, eds. Yves and Luc van Doorslaer, Amsterdam: John Benjamins, 81–104.
- Франкопуло 2013: Gil Francopoulo (ed.), *LMF. Lexical Markup Framework*, London: ISTE.
- Фреге 1892: Gottlob Frege, Über Sinn und Bedeutung, *Zeitschrift für Philosophie und philosophische Kritik* 100, 25–50.
- Шипка 2007: Danko Šipka, Ambiguous definitions: A case study, *Poznan Studies in Contemporary Linguistics* 42, 269–278.
- Шипка 2015: Danko Šipka, *Lexical Conflict: Theory and Practice*, Cambridge: Cambridge University Press.

*Danko M. Šipka*

## DETECTION OF INCONSISTENCIES IN THE TREATMENT OF MULTIPLE EQUIVALENCE: A PROLEGOMENON

### S u m m a r y

The present paper is meant to be the first step in developing the mechanisms for automatic detection of inconsistencies in lexicographic treatment of multiple equivalence, the most prominent form of lexical anisomorphism. In addition to the obvious interest this phenomenon has been commanding in lexicography (e.g., Zgusta 1971, Al-Kasimi 1977, Yong–Peng 2007), important direct and indirect contributions to the study of multiple equivalence can be found in various other intellectual traditions from ordinary language philosophy (most notably in Frege 1982 and Putnam 1975), to linguistic anthropology (Goddard 2012), translatology (Pym 2009), and second language teaching (Kramsch 1993). Šipka (2015) has identified the following common errors in the treatment of multiple equivalence in bilingual dictionaries: a. strategy mismatch, b. vagueness, c. lack of explanation or specification, d. redundancy, e. underspecification overlaps, f. overspecification gaps, g. mislabeling, h. inadequate exemplification, i. language mismatch, j. inconsist-

ent segregation, k. equivalent omission and l. overinterpretation. Automatic detection of the aforementioned errors in bilingual dictionaries assumes a feasible and reasonably accurate search pattern of some kind. A search pattern, in turn, needs to be efficient and its construction should not be overly time-consuming. More precisely, when deciding if it makes sense to develop a search pattern one should be guided by the following consideration: design feasibility, design resources complexity, comprehensiveness, and exclusiveness. The first parameter of a possible search pattern is a simple check if the error in question lends itself to automatic detection. If the answer to the first question above is positive, i.e., if the development of a search pattern is possible given the type of error, the next question should be what kinds of resources one would need to muster to design that particular search pattern. It may be so that gathering the resources for a search pattern is overly time consuming and that it makes more sense to search for the errors manually, by visually scanning the dictionary database. The next step in building the search pattern is the question about the proportion of the errors in question that will be detected by that particular search pattern. Ideally, the search pattern should detect all instances of treatment errors, but it may make sense to deploy the patterns that find most of them, especially if the design of the pattern is not overly time consuming. In this situation of underdetection, the cases that are not found by the search pattern may be addressed manually, left to be corrected in other editions of the same dictionary, etc. Finally, we want the search pattern to exclude the cases, which are not errors that we are looking for. Ideally, the output of the search patterns should not contain any non-errors (but, realistically, some level of overdetection can be tolerated). Possible search patterns for each of the aforementioned types of multiple equivalence are discussed in the central section of the present paper.

*Keywords:* multiple equivalence, lexical anisomorphism, lexicographic treatment, detection of inconsistencies, Bilingual Lexicography.